

WHAT IS CLAIMED IS:

1. A method for increasing relevance of search results, the method comprising:

receiving a search query;

5 identifying the search query as ambiguous;

clustering a plurality of search results for the search query into interpretation clusters; and

generating a search results list based on the interpretation clusters.

10 2. The method of Claim 1 wherein generating the search results list further comprises ordering search results based on the interpretation clusters.

15 3. The method of Claim 2 wherein identifying the search query as ambiguous comprises identifying at least one of: (a) an inherent ambiguity in the search query; (b) an under specified search query; or (c) a sub-phrase match for the search query.

4. The method of Claim 3 wherein identifying the search query as ambiguous comprises identifying an absence of an exact match search result with the search query.

20 5. The method of Claim 4 wherein clustering the plurality of search results associated with the search query comprises using at least one of: (a) bidden phrases, or (b) clickthrough data.

25 6. The method of Claim 1 wherein identifying the search query as ambiguous comprises identifying absence of an exact match search result with the search query.

7. The method of Claim 1 wherein clustering the plurality of search results associated with the search query comprises using at least one of: (a) bidden phrases, or (b) clickthrough data.

8. The method of Claim 7 wherein clustering the plurality of search results associated with the search query comprises using bidden phrases.

9. The method of Claim 1 wherein generating the search results list comprises interleaving search results from different interpretation clusters.

10. The method of Claim 1 comprising identifying a plurality of search results as potential matches for the search query.

11. The method of Claim 10 wherein clustering comprises using at least one of: (a) bidden search terms, or (b) searcher clickthrough terms.

12. The method of Claim 11 wherein clustering comprises calculating Pearson's correlation coefficients.

13. The method of Claim 12 wherein clustering comprises calculating an occurrence matrix.

14. The method of Claim 13 wherein identifying a plurality of search results comprises identifying bidden search terms.

15. The method of Claim 12 wherein clustering comprises

calculating a co-occurrence matrix.

16. The method of Claim 15 wherein identifying a plurality of search results comprises identifying bidden search terms.

5 17. The method of Claim 1 wherein clustering comprises:  
identifying content providers corresponding to the plurality of search results;

retrieving bidden search terms associated with each identified content provider;

10 determining correlations between the identified content providers based on the retrieved bidden search terms; and

clustering positively correlating content providers based on the correlations.

18. The method of Claim 17 wherein clustering positively 15 correlating content providers comprises:

summing correlation scores of positively correlating content providers for each content provider;

selecting a highest scored content provider;

20 forming a cluster including content providers having positive correlation with the highest scored content provider; and

repeating summing, selecting, and forming after removing previously clustered content providers.

19. The method of Claim 18 wherein clustering comprises 25 using correlations of occurrence values.

20. The method of Claim 19 wherein clustering comprises calculating Pearson's correlation coefficients.

21. The method of Claim 18 wherein clustering comprises using correlations of co-occurrence values.

22. The method of Claim 21 wherein clustering comprises calculating Pearson's correlation coefficients.

5 23. The method of Claim 17 wherein clustering positively correlating content providers comprises:

evaluating pairs of content providers having positive correlation score in descending order of correlation score;

10 assigning content providers to clusters after each evaluation of a pair content providers, wherein assigning comprises:

15 creating a new cluster comprising each content provider of a pair of content providers if neither content provider is assigned to a cluster and assigning each content provider of a pair of content providers to the new cluster; and

20 assigning an unassigned content provider of a partially assigned pair of content providers to a cluster comprising a previously assigned content provider of the partially assigned pair of content providers if only one of the content providers of a pair of content providers was previously assigned to a cluster.

24. The method of Claim 23 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously assigned if both content providers of a pair are already assigned.

25. The method of Claim 23 further comprising clustering clusters so as to further define a hierarchical structure between

the content providers.

26. The method of Claim 25 further comprising storing relationship information corresponding to the hierarchical structure of clustered clusters.

5 27. The method of Claim 23 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

clustering clusters if both content providers of a pair are already assigned to clusters comprising:

10 defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

incrementing an inter-cluster edge score by the correlation score associated with the pair of content providers.

15 28. The method of Claim 27 further comprising normalizing the incremented inter-cluster score.

20 29. The method of Claim 28 wherein normalizing comprises dividing the incremented inter-cluster score by a product of a number of objects in each of the clusters of previously assigned content providers.

30. The method of Claim 23 further comprising sorting from maximum to minimum pairs of content providers having positive correlation score prior to evaluating.

31. A method for disambiguation of search phrases, the method comprising:

identifying interpretation clusters using bidden search terms of content providers;

5 influencing a position of a search result in a search results list based on the interpretation clusters; and

providing the search results list in response to a search query received by a searcher for review by the searcher.

32. The method of Claim 31 wherein identifying interpretation clusters comprises using correlations of occurrence values.

10 33. The method of Claim 32 wherein identifying interpretation clusters comprises calculating Pearson's correlation coefficients and clustering along positive correlations.

15 34. The method of Claim 33 wherein clustering comprises:

identifying content providers corresponding to the plurality of search results;

20 retrieving bidden search terms associated with each identified content provider;

determining correlations between the identified content providers based on the retrieved bidden search terms; and

clustering positively correlating content providers based on the correlations.

25 35. The method of Claim 34 wherein clustering positively correlating content providers comprises:

summing correlation scores of positively correlating content providers for each content provider;

selecting a highest scored content provider;  
forming a cluster including content providers having  
positive correlation with the highest scored content provider;  
and

5 repeating summing, selecting, and forming after  
removing previously clustered content providers.

36. The method of Claim 31 wherein clustering comprises  
using correlations of co-occurrence values.

10 37. The method of Claim 36 wherein identifying  
interpretation clusters comprises calculating Pearson's  
correlation coefficients and clustering along positive  
correlations.

38. The method of Claim 36 wherein identifying  
interpretation clusters comprises:

15 calculating a matrix of bidden terms in common;  
calculating correlation coefficients corresponding to  
the bidden terms in common; and  
assigning interpretation clusters using the correlation  
coefficients.

20 39. The method of Claim 38 wherein calculating correlation  
coefficients comprises calculating Pearson's correlation  
coefficients and clustering along positive correlations.

40. The method of Claim 39 wherein clustering comprises:  
identifying content providers corresponding to the  
25 plurality of search results;  
retrieving bidden search terms associated with each  
identified content provider;

determining correlations between the identified content providers based on the retrieved bidden search terms; and

clustering positively correlating content providers based on the correlations.

5        41. The method of Claim 40 wherein clustering positively correlating content providers comprises:

      summing correlation scores of positively correlating content providers for each content provider;

      selecting a highest scored content provider;

10        forming a cluster including content providers having positive correlation with the highest scored content provider; and

      repeating summing, selecting, and forming after removing previously clustered content providers.

15        42. The method of Claim 31 wherein identifying interpretation clusters comprises:

      evaluating pairs of content providers having positive correlation coefficients in descending order; and

20        assigning content providers to clusters after each evaluation of a pair content providers, wherein assigning comprises:

      creating a new cluster comprising each content provider of a pair of content providers if neither content provider is assigned to a cluster and assigning each content provider of a pair of content providers to the new cluster; and

25        assigning an unassigned content provider of a partially assigned pair of content providers to a cluster comprising a previously assigned content provider of the partially assigned pair of content providers if only one of the content providers of a pair of content providers was previously

assigned to a cluster.

43. The method of Claim 42 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously 5 assigned if both content providers of a pair are already assigned.

44. The method of Claim 43 further comprising clustering clusters so as to further define a hierarchical structure between the content providers.

10 45. The method of Claim 44 further comprising storing relationship information corresponding to the hierarchical structure of clustered clusters.

15 46. The method of Claim 42 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

clustering clusters if both content providers of a pair are already assigned to clusters comprising:

20 defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

incrementing an inter-cluster edge score by the correlation coefficient associated with the pair of content providers.

25 47. The method of Claim 46 further comprising normalizing the incremented inter-cluster score.

48. The method of Claim 42 further comprising sorting from

maximum to minimum pairs of content providers having positive correlation coefficient prior to evaluating.

49. The method of Claim 31 wherein identifying interpretation clusters comprises evaluating redundant term  
5 vectors.

50. The method of Claim 31 wherein identifying interpretation clusters comprises defining a cluster signature based on an evaluation of probabilities of content providers being members of clusters.

10 51. The method of Claim 31 wherein providing the search results comprises providing search results capable of review by a searcher wherein the searcher is an automated process representing a user's interest.

52. A method for interpretation clustering, the method comprising:

identifying a plurality of terms and associated content providers;

5 calculating a matrix from the identified plurality of terms and associated content providers;

calculating similarity scores between content providers based on the matrix; and

10 assigning interpretation clusters using the similarity scores.

53. The method of Claim 52 wherein calculating similarity scores comprises calculating Pearson's correlation coefficients.

54. The method of Claim 53 wherein calculating a matrix comprises calculating an occurrence matrix.

15 55. The method of Claim 54 wherein identifying comprises using bidden search terms.

56. The method of Claim 54 wherein identifying comprises using searcher clickthrough terms.

20 57. The method of Claim 53 wherein calculating a matrix comprises calculating a co-occurrence matrix.

58. The method of Claim 57 wherein identifying comprises using at least one of: (a) bidden search terms, or (b) searcher clickthrough terms.

25 59. The method of Claim 52 wherein calculating a matrix comprises calculating a co-occurrence matrix.

60. The method of Claim 59 wherein identifying comprises using bidden search terms.

61. The method of Claim 59 wherein identifying comprises using searcher clickthrough terms.

5

62. The method of Claim 52 wherein assigning interpretation clusters comprises:

evaluating pairs of content providers having correlating similarity scores in descending order of similarity score; and

assigning content providers to clusters after each evaluation of a pair content providers, wherein assigning comprises:

10 creating a new cluster comprising each content provider of a pair of content providers if neither content provider is assigned to a cluster and assigning each content provider of a pair of content providers to the new cluster; and

15 assigning an unassigned content provider of a partially assigned pair of content providers to a cluster comprising a previously assigned content provider of the partially assigned pair of content providers if only one of the content providers of a pair of content providers was previously assigned to a cluster.

20 25 63. The method of Claim 62 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously assigned if both content providers of a pair are already assigned.

64. The method of Claim 62 further comprising clustering clusters to define a hierarchical structure between the content providers.

5 65. The method of Claim 64 further comprising storing relationship information corresponding to the hierarchical structure of clustered clusters.

66. The method of Claim 62 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

10 clustering the clusters if both content providers of a pair are already assigned comprising:

defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

15 incrementing an inter-cluster edge score by the similarity score associated with the pair of content providers.

67. The method of Claim 66 further comprising normalizing the incremented inter-cluster score.

20 68. The method of Claim 67 wherein normalizing comprises dividing the incremented inter-cluster score by a product of a number of objects in each of the clusters of previously assigned content providers.

25 69. The method of Claim 62 further comprising sorting from maximum to minimum pairs of content providers having correlating similarity score prior to evaluating.

70. A method for interpretation clustering, the method comprising:

identifying a plurality of terms and associated content providers;

5 calculating a co-occurrence matrix of terms in common; calculating correlation coefficients corresponding to the terms in common; and

assigning interpretation clusters using the correlation coefficients.

10 71. The method of Claim 70 wherein identifying the plurality of terms comprises identifying one of: (a) bidden search terms, or (b) click-through terms.

15 72. The method of Claim 71 wherein calculating correlation coefficients comprises calculating normalized correlation coefficients.

73. The method of Claim 72 wherein assigning interpretation clusters comprises clustering along positive correlations.

20 74. The method of Claim 73 wherein clustering positively correlating content providers comprises:

summing correlation scores of positively correlating content providers for each content provider;

selecting a highest scored content provider;

25 forming a cluster including content providers having positive correlation with the highest scored content provider; and

repeating summing, selecting, and forming after removing previously clustered content providers.

75. The method of Claim 71 wherein identifying interpretation clusters comprises:

evaluating pairs of primary entities having positive correlation coefficients in descending order; and

5 assigning primary entities to clusters after each evaluation of a pair primary entities, wherein assigning comprises:

10 creating a new cluster comprising each content provider of a pair of primary entities if neither content provider is assigned to a cluster and assigning each content provider of a pair of primary entities to the new cluster; and

15 assigning an unassigned content provider of a partially assigned pair of primary entities to a cluster comprising a previously assigned content provider of the partially assigned pair of primary entities if only one of the content providers of a pair of content providers was previously assigned to a cluster.

20 76. The method of Claim 75 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously assigned if both content providers of a pair are already assigned.

25 77. The method of Claim 76 further comprising clustering clusters so as to further define a hierarchical structure between the content providers.

78. The method of Claim 77 further comprising storing relationship information corresponding to the hierarchical structure of clustered clusters.

79. The method of Claim 75 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

5 clustering clusters if both content providers of a pair are already assigned to clusters comprising:

defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

10 incrementing an inter-cluster edge score by the correlation coefficient associated with the pair of content providers.

80. The method of Claim 79 further comprising normalizing the incremented inter-cluster score.

15 81. The method of Claim 75 further comprising sorting from maximum to minimum pairs of content providers having positive correlation coefficient prior to evaluating.

82. The method of Claim 70 wherein identifying the plurality of terms comprises selecting search terms preselected by a content provider.

20 83. The method of Claim 82 wherein identifying the plurality of terms comprises selecting search terms preapproved for relevance with the content provider.

84. The method of Claim 82 wherein identifying the plurality of terms comprises selecting bidden search terms.

25 85. The method of Claim 70 wherein identifying the plurality of terms and associated content providers comprises

including search terms of a content provider representative of a directory category.

86. A method for a database search system, the method comprising:

identifying interpretation clusters based on shared bidden search terms of content providers; and

5 using the interpretation clusters for one of: (a) identifying a cluster including a representative content provider; (b) evaluating search terms for canonicalization; (c) evaluating a selection of a search term by a content provider; (d) assigning an ambiguity score to selection of a search term by 10 a content provider; (e) ordering of search results; or (f) selecting a bidden search result for a search query without an exact match.

87. The method of Claim 86 wherein using the interpretation clusters comprises identifying a cluster including a representative content provider, and wherein identifying interpretation clusters comprises including bidden terms from a content provider representative of a directory category and selecting an interpretation cluster containing the content provider representative of the directory category.

20 88. The method of Claim 86 wherein using the interpretation clusters comprises identifying a cluster including a representative content provider, and further comprising selecting a search result from the cluster including the representative content provider.

25 89. The method of Claim 86 wherein identifying interpretation clusters comprises:

identifying bidden search terms and associated content providers;

calculating a matrix of bidden search terms in common

between the content providers;

calculating correlation coefficients corresponding to the bidden search terms in common; and

5 assigning interpretation clusters using the correlation coefficients.

90. The method of Claim 86 wherein using the interpretation clusters comprises evaluating search terms for canonicalization, and further comprising identifying search terms for mapping to a same canonical form.

10 91. The method of Claim 86 wherein using the interpretation clusters comprises evaluating search terms for canonicalization, and further comprising identifying search terms not suitable for mapping to a same canonical form.

15 92. The method of Claim 86 wherein using the interpretation clusters comprises evaluating a selection of a search term by a content provider, and further comprising generating an automated response regarding approval of a search term selection by a content provider based on a comparison of interpretation clusters of the content provider making the search term selection with 20 interpretation clusters of other content providers already associated with the search term selection.

25 93. The method of Claim 86 wherein using the interpretation clusters comprises assigning an ambiguity score to selection of a search term by a content provider, and further comprising prioritizing an editorial effort based on the assigned ambiguity score.

94. The method of Claim 93 further comprising prioritizing

hand labeling of ambiguous search phrases based on the ambiguity score.

95. The method of Claim 86 wherein using the interpretation clusters comprises selecting a bidden search result for a search query without an exact match, and further comprising identifying bidden search term matches to sub-phrases of the search query.

96. The method of Claim 86 wherein using the interpretation clusters comprises using the interpretation clusters for automated evaluation of a selection of a search term by a content provider.

97. A method for producing search results consistent with a directory category in a database search system, the method comprising:

5 receiving a search query having an identified directory category;

identifying a plurality of bidden search results as potential matches for the search query;

10 including bidden terms from a content provider representative of the identified directory category in the plurality of search results;

clustering the bidden search results including the bidden terms from the content provider representative of the identified directory category; and

15 generating category consistent search results from the cluster comprising the provider representative of the identified directory category.

98. The method of Claim 97 wherein clustering comprises calculating Pearson's correlation coefficients.

20 99. The method of Claim 98 wherein clustering comprises identifying content providers having positive correlation with the content provider representative of the identified directory category.

25 100. The method of Claim 98 wherein clustering comprises using an occurrence matrix of content providers and bidden search terms.

101. The method of Claim 98 wherein clustering comprises using a terms-in-common co-occurrence matrix.

102. A computer readable medium comprising executable instructions for performing a method for providing search results in response to a search query, the method comprising:

5                   identifying the search query as ambiguous;

                  clustering a plurality of search results for the search query into interpretation clusters using at least one of (a) bidden phrases, or (b) clickthrough data; and

                  generating a search results list comprising ordering results based on the interpretation clusters.

10                103. The computer readable medium of Claim 102 wherein clustering comprises:

                  identifying content providers corresponding to the plurality of search results;

15                retrieving bidden search terms associated with each identified content provider;

                  determining correlations between the identified content providers based on the retrieved bidden search terms; and

                  clustering positively correlating content providers based on the correlations.

20                104. The computer readable medium of Claim 104 wherein clustering positively correlating content providers comprises:

                  summing correlation scores of positively correlating content providers for each content provider;

25                selecting a highest scored content provider;

                  forming a cluster including content providers having positive correlation with the highest scored content provider; and

                  repeating summing, selecting, and forming after removing previously clustered content providers.

105. The computer readable medium of Claim 105 wherein clustering comprises using correlations of occurrence values.

106. The computer readable medium of Claim 105 wherein clustering comprises using correlations of co-occurrence values.

5 107. The computer readable medium of Claim 104 wherein clustering positively correlating content providers comprises:

evaluating pairs of content providers having positive correlation score in descending order of correlation score; and

10 assigning content providers to clusters after each evaluation of a pair content providers, wherein assigning comprises:

15 creating a new cluster comprising each content provider of a pair of content providers if neither content provider is assigned to a cluster and assigning each content provider of a pair of content providers to the new cluster; and

20 assigning an unassigned content provider of a partially assigned pair of content providers to a cluster comprising a previously assigned content provider of the partially assigned pair of content providers if only one of the content providers of a pair of content providers was previously assigned to a cluster.

25 108. The computer readable medium of Claim 110 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously assigned if both content providers of a pair are already assigned.

109. The computer readable medium of Claim 110 further comprising clustering clusters so as to further define a

hierarchical structure between the content providers.

110. The computer readable medium of Claim 110 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

5 clustering clusters if both content providers of a pair are already assigned to clusters comprising:

defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

10 incrementing an inter-cluster edge score by the correlation score associated with the pair of content providers.

111. The computer readable medium of Claim 114 further comprising normalizing the incremented inter-cluster score.

112. A system for providing search results in response to a search query, the system comprising:

a computer adapted to receive a search query;

a processor associated with the computer, the processor

5 being programmed for:

identifying the search query as ambiguous;

clustering a plurality of search results for the search query into interpretation clusters using at least one of (a) bidden phrases, or (b) clickthrough data; and

10 generating a search results list comprising ordering results based on the interpretation clusters.

113. The system of Claim 102 wherein clustering comprises:

identifying content providers corresponding to the plurality of search results;

15 retrieving bidden search terms associated with each identified content provider;

determining correlations between the identified content providers based on the retrieved bidden search terms; and

20 clustering positively correlating content providers based on the correlations.

114. The system of Claim 104 wherein clustering positively correlating content providers comprises:

summing correlation scores of positively correlating content providers for each content provider;

25 selecting a highest scored content provider;

forming a cluster including content providers having positive correlation with the highest scored content provider; and

30 repeating summing, selecting, and forming after removing previously clustered content providers.

115. The system of Claim 105 wherein clustering comprises using correlations of occurrence values.

116. The system of Claim 105 wherein clustering comprises using correlations of co-occurrence values.

5 117. The system of Claim 104 wherein clustering positively correlating content providers comprises:

evaluating pairs of content providers having positive correlation score in descending order of correlation score; and

10 assigning content providers to clusters after each evaluation of a pair content providers, wherein assigning comprises:

15 creating a new cluster comprising each content provider of a pair of content providers if neither content provider is assigned to a cluster and assigning each content provider of a pair of content providers to the new cluster; and

20 assigning an unassigned content provider of a partially assigned pair of content providers to a cluster comprising a previously assigned content provider of the partially assigned pair of content providers if only one of the content providers of a pair of content providers was previously assigned to a cluster.

25 118. The system of Claim 110 wherein assigning content providers to clusters after each evaluation of a pair content providers comprises leaving content providers as previously assigned if both content providers of a pair are already assigned.

119. The system of Claim 110 further comprising clustering

clusters so as to further define a hierarchical structure between the content providers.

120. The system of Claim 110 wherein assigning content providers to clusters after each evaluation of a pair content providers further comprises:

clustering clusters if both content providers of a pair are already assigned to clusters comprising:

defining an inter-cluster score between the clusters of previously assigned content providers of a pair of content providers; and

incrementing an inter-cluster edge score by the correlation score associated with the pair of content providers.

121. The system of Claim 114 further comprising normalizing the incremented inter-cluster score.